

Application No. 09/692,846
Filed 10/19/00
Page 20

The specification and claims have been amended to correct minor informalities and to address other issues raised by the Examiner. Claims 26 through 56 have been added. Claims 1, 5, 6, 8, 13, 15, 16, and 17 have been amended. Claims 18 through 25 have been canceled. Forty-eight claims remain pending in the application: Claims 1 through 17 and 26 through 56. Reconsideration of Claims 1 through 17 and consideration of claims 26 through 56 in view of the amendments above and arguments below is respectfully requested.

By way of this amendment, Applicants have made a diligent effort to place the claims in condition for allowance. However, should there remain any outstanding issues that require adverse action, it is respectfully requested that Examiner telephone the undersigned at (858) 587-7644 so that such issues may be resolved as expeditiously as possible.

Turning to the specific objections and rejections:

1. Claims 1 through 4, 6 through 10, and 18 through 25 stand rejected under 35 U.S.C. § 102(b), as being anticipated

Application No. 09/692,846
Filed 10/19/00
Page 21

by Wyard, et al., "Spoken Language Systems"--beyond prompt and response" (Wyard, et al.). Claims 18 through 25 have been cancelled herewith. Thus, the rejection of Claims 18 through 25 is now moot.

Claim 1 has been amended herewith so as to clarify that the recited microphone array is a *three-dimensional microphone array*. As is noted on page 5, numbered paragraph 6, et seq., of the present office action, Wyard et al. *does not* teach a three-dimensional microphone array. Accordingly, the rejection under 35 U.S.C. § 102(b) is overcome.

Furthermore, with regard to the rejection of Claim 5, which, as originally filed, recited a three-dimensional microphone array, the Examiner states "this feature is well known in the art." No support is provided for this conclusion. Consistent with this, Applicants assert that the application of a three-dimensional microphone array to the specific structure claimed in Claim 1 is far from known or obvious. Nothing in the Wyard, et al. reference would suggest to one of ordinary skill in the art that a three-dimensional microphone array be employed. In fact, Wyard, et al. do not appear to recognize a

Application No. 09/692,846
Filed 10/19/00
Page 22

need for or any advantage of anything other than a single microphone.

As applicants claim a three-dimensional microphone array (or as originally stated, a first microphone array), this rejection is overcome.

If the Examiner is relying on personal knowledge as a basis for the assertion that the three-dimensional microphone array is well known (which in and of itself is, of course, insufficient to sustain the *prima facie* case of obviousness), and furthermore for an assertion that the prior art teaches or suggests the use of such a three-dimensional microphone array in a structure such as recited in Claim 1, Applicants respectfully request that the Examiner submit into the record an affidavit to this effect. See M.P.E.P. § 2144.03. Absent such a showing, the present rejection is in error.

Therefore reconsideration and withdrawal of such rejection is respectfully requested.

Claims 2 through 16 are dependent on Claim 1 and thus their rejection is similarly overcome for at least those reasons stated above.

Application No. 09/692,846
Filed 10/19/00
Page 23

Specifically with regard to Claim 2, it is asserted that Wyard et al. describe a plurality of devices coupled to a natural language interface. While Wyard, et al. describe several applications for natural language systems, at no time do they describe multiple devices simultaneously coupled to a device interface and operated by the natural language interface, as claimed by applicants.

It is to be emphasized that while Wyard et al., in general, describe a number of different and discrete systems and examples, it is improper for these discrete systems and examples to be combined, thus deviating from what is described within the four corners of the Wyard et al. reference, in order to build a device that complies with the limitations of Applicants' claims. To do so is inconsistent with the requirements of a rejection under 35 U.S.C. § 102(b).

Accordingly, the present rejection of Claim 2 is overcome of these additional reasons.

With specific regard to Claim 6, at no point does Wyard, et al. describe multiple devices controlled by a natural language interface in a single system. Accordingly, it is

Application No. 09/692,846
Filed 10/19/00
Page 24

impossible that Wyard et al. describe the extraction of each of a plurality of devices into a respective one or a plurality of grammars and a respective one of a plurality of lexicons corresponding to each of the plurality of devices. Accordingly, Claim 6 is not taught or suggested by the Wyard et al. reference and the present rejection is therefore overcome on this additional basis.

With specific regard to Claim 7, attention is drawn to page 187 of the Wyard et al. reference wherein it states "moving beyond the menu-style interaction toward conversational spoken language will allow users to express the requirements more directly and avoid tedious navigation through menus." This language is cited in the present office action as a basis for teaching the claimed feature in which the "natural language interface module searches for the non-prompted, open-ended user requests upon the receipt and recognition of an attention word." This feature is simply not taught by the cited language, nor is such feature taught elsewhere in the Wyard, et al. reference, and thus the rejection of Claim 7 is overcome for this additional reason.

Application No. 09/692,846
Filed 10/19/00
Page 25

With more specific regard to the rejection of Claim 8, again Wyard, et al. does not teach the use of an attention word and thus cannot possibly teach the context switching of grammars, acoustic models, and lexicons upon the receipt and recognition of an attention word. Accordingly, for this additional reason, the rejection of Claim 8 is overcome.

With more specific regard to Claims 9 and 10, the recitation of a "grammar module for storing different grammars for each of the plurality of devices" or "an acoustic model module for storing different acoustic models for each of the plurality of devices" is recited. At no time does Wyard et al. teach the controlling of multiple devices simultaneously. Thus, Wyard et al. cannot possibly teach the grammar module for storing different grammars for each of the plurality of devices or the acoustic model module for storing different acoustic models for each of the plurality of devices.

2. Claim 17 stands rejected under 35 U.S.C. § 102(b) as being anticipated by United States Patent No. 5,797,123 (Chow, et al.).

Application No. 09/692,846
Filed 10/19/00
Page 26

Chow et al. are cited as teaching a method of speech recognition comprising the two steps of first searching for an attention word based on first context including a first set of models, grammars, and lexicons and then switching upon finding the attention word to a second context including a second set of models, grammars, and lexicons. While citation is made to FIG. 1 in column 5, lines 50 through 57, and column 6, lines 1 through 67, nowhere in this cited portion of the Chow et al. patent document (or elsewhere) is a description made of first using a first context to search for an attention word and then, upon finding the attention word, switching to a second context.

What is described by Chow, et al. is the use of key phrases selected by a key-phrase detector on the basis of key-phrase grammars, and the assembly of these key phrases into sentences using a sentence hypothesizer. Such functionality is carried out using a single set of models, grammars, and lexicons, and at no point is switching to a second context, i.e., a second set of models, grammars, and lexicons, described.

Accordingly, Chow et al. fails to teach or even suggest that which is claimed in Claim 17, and thus the present

Application No. 09/692,846
Filed 10/19/00
Page 27

rejection is in error. Accordingly, reconsideration and withdrawal of the present rejection is respectfully requested.

3. Claims 5 and 11 through 16 stand rejected under 35 U.S.C. § 103(a) as being unpatentable over Wyard, et al.

At the outset, attention is drawn to the numbered paragraph 1, et seq., above at which it is pointed out that nowhere in the Wyard, et al reference is a referral to a three-dimensional microphone array made. Furthermore, attention is drawn to the request hereinabove that if the Examiner is relying on personal knowledge for his conclusion that such a three-dimensional arrays are known in the art, and furthermore that a combination of a three-dimensional array in combination with Applicants' claimed structure is suggested by such art, the Examiner is respectfully requested to submit an affidavit to this effect. See M.P.E.P. § 2144.03. Absent this affidavit, this rejection is overcome and reconsideration and withdrawal is requested.

As claims 11 through 16 are dependent on Claim 1, which now more clearly recites a three-dimensional microphone

Application No. 09/692,846
Filed 10/19/00
Page 28

array, Claims 11 through 16 should now be in condition for allowance on at least this basis.

Furthermore, with regard to Claim 11, the Examiner asserts that the use of a wireless device interface with a natural language interface control system for operating a plurality of devices is well known in the art. No authority for this conclusion is provided, and thus, the Examiner is respectfully requested to submit an affidavit indicating the he is relying on his personal knowledge of wireless device interfaces for coupling a natural language interface control system to a plurality of controlled devices, and that such wireless device interfaces and their combination with a natural language interface control system are well known. See M.P.E.P. § 2144.03. Absent such an affidavit, the rejection of Claim 11 is overcome.

With further specific regard to Claim 12, no showing or assertion is made in the office action and no showing is made in the art of record that a natural language interface control system could be coupled to an external network as claimed. Accordingly, as no basis for rejection is provided, Claim 12

Application No. 09/692,846
Filed 10/19/00
Page 29

appears to be allowable, and its rejection appears to be in error.

With specific regard to Claim 13, no showing or assertion is made in the office action, and no showing is made in the art of record of the use of a remote unit comprising a first microphone array, a feature extraction module, the speech recognizer module and the natural language interface module, wherein the three-dimensional microphone array comprises the first microphone array. Accordingly, as no basis for rejection is provided, Claim 13 appears allowable, and its rejection appears to be in error.

With specific regard to Claim 14, the use of the base unit coupled to a remote unit is not shown or suggested by any reference of record, nor is any assertion that such is shown or suggested by any reference of record made in the present office action. Accordingly, as no basis for rejection is provided, Claim 13 appears allowable.

With specific regard to Claim 15, no teaching or suggestion is made in the art of record that a base unit includes a second microphone array, wherein the three-

Application No. 09/692,846
Filed 10/19/00
Page 30

dimensional microphone array comprises the second microphone array. Furthermore, no assertion is made in the present office action that such is shown or suggested by any reference of record. Accordingly, as no basis for rejection is provided, Claim 15 appears to be allowable.

With respect to Claim 16, nowhere in any reference of record is a first microphone array and a second microphone array used to implement a three-dimensional microphone array. Furthermore, no assertion is made in the present office action that such is shown or suggested in any reference of record. Accordingly, Claim 16 appears, as no basis for rejection is provided, to be allowable.

Newly submitted Claims 26 through 56 are believed to be allowable because they are directed to subject matter that which is not shown or suggested in the prior art. In particular, Claims 26 through 56 each recite an external network that is employed by or coupled to natural language control system. See also above Applicants remarks regarding Claim 12, which also recites an external network. As the claimed structures and steps are not shown or suggested in any reference

Application No. 09/692,846
Filed 10/19/00
Page 31

of record, and no assertion is made in the office action that such structure or steps (such as in, e.g., Claim 12, which has already been searched and examined) are shown or suggested in the art of record, all of the new claims should be allowable.

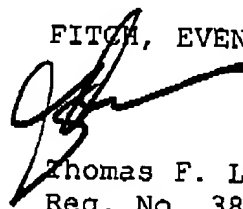
In view of the above, Applicants submit that Claims 1 through 17 and 26 through 56 are now in condition for allowance, and prompt and favorable action is earnestly solicited.

Attached hereto is a marked-up version of the changes made to the specification and claims by the current amendment. The attached pages are captioned "Version with markings to show changes made."

Please charge any deficiency or overpayment in fees to Deposit Account 06-1135.

Respectfully submitted,

FITCH, EVEN, TABIN & FLANNERY



Thomas F. Lebens
Reg. No. 38,221

Application No. 09/692,846
Filed 10/19/00
Page 32

Address all correspondence to:
FITCH, EVEN, TABIN & FLANNERY
120 So. LaSalle Street, Suite 1600
Chicago, Illinois 60603

Direct telephone inquiries to:
Thomas F. Lebens
(858) 587-7644
La Jolla, California Office of
FITCH, EVEN, TABIN & FLANNERY

Application No. 09/692,846
Filed 10/19/00
Page 33

VERSION WITH MARKINGS TO SHOW CHANGES MADE

In the specification:

Paragraph beginning at line 19, page 15, has been amended as follows:

The feature extraction module 202 performs edge-detection, signal conditioning and feature extraction. According to one embodiment, speech edge detection is accomplished using noise estimation and energy detection based on the 0th Cepstral coefficient and zero-crossing statistics. Feature extraction and signal conditioning consist of extracting Mel-frequency cepstral coefficients (MFCC), delta information and acceleration information. It is a 38 dimensional feature vector based on 12.8 ms sample buffers overlapped by 50%. Such feature extraction modules 202 and functionality are well understood in the art, and [that] one skilled in the art may implement the feature extraction module in a variety of ways. Thus, the output of the feature extraction module 202 is a sequence of feature vectors.

Application No. 09/692,846
Filed 10/19/00
Page 34

Paragraph beginning at line 19, page 26, has been amended as follows:

In some embodiments, a caching scheme is used for the lexicons stored in memory on the remote unit, e.g., by the N-gram grammar module 218. As stated above, a lexicon is a dictionary consisting of words and their pronunciation entries. These pronunciations may be implemented as either phonetic spellings that refer to phonetic models, or to whole-word models. A given word entry may contain alternate pronunciation entries, most of which are seldom used by any single speaker. This redundancy is echoed at each part-of-speech abstraction, creating even more entries that are never utilized by a given speaker. This implies that if lexicon entries are sorted by their frequency of usage, there is a great chance that the words in an utterance can be found among the top n lexicon entries. As such, the cache is divided into different levels divided by frequency of use. For example, frequently used lexicon entries will be stored within the top level of the cache. A caching scheme may be devised in which the top 10% of the cache is used

Application No. 09/692,846
Filed 10/19/00
Page 35

90% of the time, for example. Thus, according to an embodiment, a multi-pass search is performed where the most likely entries are considered in the first pass. If the garbage score from this pass is high enough to believe that the words actually spoken were contained in the set of most likely spellings, the speech decoder 216 reports the results to the calling function. If this score is low, the system falls back to considering a wider range of spellings. If the score from the first pass is high, but not high enough in order to be able to make a decision whether the correct spellings, for the elements of the utterance, were contained in the set of most likely spellings, this is also reported back to the calling function, which might prompt the user for clarification. If a lexicon spelling for a given part-of-speech is never used while some of its alternative spellings are frequently used, that spelling is put in a "trash can" and will never be considered for that user. As such, rarely used spellings are not considered and the chance of confusing similar-sounding utterances with one of those spellings is reduced and the recognition accuracy is therefore

Application No. 09/692,846
Filed 10/19/00
Page 36

increased. Further, the caching scheme allows the system to consider less data and hence provides a great speed improvement.

Paragraph beginning at line 1, page 35, has been amended as follows:

Furthermore, the NLICS 102 may download command signals for the device abstraction module of the remote unit 104. For example, a user would like to operate an older VCR that has an IR remote control manufactured by a different maker [that] than the NLICS. The base unit 106 simply downloads the commands that are stored for any number of devices. These commands are then stored in the device abstraction module. Also, the NLICS can submit feature vector data and labels associated with high-confidence utterances to the collaborative corpus. This data can then be incorporated with other data and used to train improved models that are subsequently redistributed. This approach can also be used to incorporate new words into the collaborative corpus by submitting the feature vector data and its label, which may subsequently be combined with other data and phonetically transcribed using the forward-backward

Application No. 09/692,846
Filed 10/19/00
Page 37

algorithm. This entry may then be added to the lexicon and redistributed.

Paragraph beginning at line 3, page 41, has been amended as follows:

If the natural language interface module 222 detects a user request (Step 416), i.e. a user request (as determined by the PCFG grammar system and device context) is received, it draws [on] one of three conclusions (Steps 420, 422 or 424). According to Step 420, the user request is unambiguously understood and the natural language interface module can comply with a user request. Thus, the natural language interface module 222 carries out the command by sending the appropriate signals via the device interface 210, as indicated by the device abstraction. Then, the context of the speech recognition module 204 and the natural language interface module 206 is switched back to look for attention words (Step 426), before proceeding to Step 404.

In the claims:

Application No. 09/692,846
Filed 10/19/00
Page 38

Claim 1 has been amended as follows:

SUB A) 1. (Amended) A natural language interface control system for operating a plurality of devices comprising:

- a 3 dimensional ~~(first)~~ microphone array;
- a feature extraction module coupled to the first microphone array;
- a speech recognition module coupled to the feature extraction module, wherein the speech recognition module utilizes hidden Markov models;
- a natural language interface module coupled to the speech recognition module; and
- a device interface coupled to the natural language interface module, wherein the natural language interface module is for operating a plurality of devices coupled to the device interface based upon non-prompted, open-ended natural language requests from a user.

Claim 5 has been amended as follows:

5. (Amended) The system of Claim 1 wherein the microphone array comprises said ~~a~~ 3 dimensional microphone

Application No. 09/692,846
Filed 10/19/00
Page 39

array further comprising a planar microphone array and at least one linear microphone array located in a different plane in space.

Claim 6 has been amended as follows:

Sub A2 6. (Amended) The system of Claim 1 wherein the natural language interface abstracts each of the plurality of devices into a respective one of a plurality of grammars and a respective one of a plurality of [lexicons] lexica corresponding to each of the plurality of devices.

Claim 8 has been amended as follows:

8. (Amended) The system of Claim 1 wherein the natural language interface module context switches grammars, acoustic models, and [lexicons] lexica upon receipt and recognition of an attention word.

Claim 13 has been amended as follows:

13. The system of Claim 1 further comprising a remote unit containing a[the] first microphone array, the feature

Application No. 09/692,846
Filed 10/19/00
Page 40

extraction module, the speech recognition module, and the natural language interface module, wherein said 3 dimensional microphone array includes the first microphone array.

Claim 15 has been amended as follows:

15. (Amended) The system of Claim 14 wherein the base unit includes a second microphone array, wherein said 3 dimensional microphone array includes the second microphone array.

Claim 16 has been amended as follows:

16. (Amended) The system of Claim 15 wherein the first microphone array and the second microphone array implement said[a] 3 dimensional microphone array.

Claim 17 has been amended as follows:

17. (Amended) A method of speech recognition comprising:

searching for an attention word based on a first context including a first set of models, grammars, and

Application No. 09/692,846
Filed 10/19/00
Page 41

[lexicons] lexica; and

switching, upon finding the attention word, to a second context to search for an open-ended user request, wherein second context includes a second set of models, grammars, and lexicons.

New claims have been added as follows:

26. (new) A natural language interface control system for operating a plurality of devices comprising:

a first microphone;

a feature extraction module coupled to the first microphone;

a speech recognition module coupled to the feature extraction module;

a natural language interface module coupled to the speech recognition module;

a device interface coupled to the natural language interface module, wherein the natural language interface module is for operating a plurality of devices coupled to the device

Application No. 09/692,846
Filed 10/19/00
Page 42

interface based upon non-prompted, open-ended natural language requests from a user; and

an external network interface coupled to the natural language interface control system.

27. (new) The system of Claim 26 further comprising the plurality of devices coupled to the natural language interface module.

28. (new) The system of Claim 26 wherein the speech recognition module utilizes an N gram grammar.

29. (new) The system of Claim 26 wherein the natural language interface module utilizes a probabilistic context free grammar.

30. (new) The system of Claim 26 wherein the microphone array comprises a 3 dimensional microphone array comprising a planar microphone array and at least one linear microphone array located in a different plane in space.

Application No. 09/692,846
Filed 10/19/00
Page 43

31. (new) The system of Claim 26 wherein the natural language interface abstracts each of the plurality of devices into a respective one of a plurality of grammars and a respective one of a plurality of lexica corresponding to each of the plurality of devices.

32. (new) The system of Claim 26 wherein the natural language interface module searches for the non-prompted, open-ended user requests upon the receipt and recognition of an attention word.

33. (new) The system of Claim 26 wherein the natural language interface module context switches grammars, acoustic models, and lexica upon receipt and recognition of an attention word.

34. (new) The system of Claim 26 further comprising a grammar module for storing different grammars for each of the plurality of devices.

Application No. 09/692,846
Filed 10/19/00
Page 44

35. (new) The system of Claim 26 further comprising an acoustic model module for storing different acoustic models for each of the plurality of devices.

36. (new) The system of Claim 26 wherein the device interface comprises a wireless device interface.

37. (new) The system of Claim 26 further comprising a remote unit containing the first microphone array, the feature extraction module, the speech recognition module, and the natural language interface module.

38. (new) The system of Claim 37 further comprising a base unit coupled to the remote unit.

39. (new) The system of Claim 38 wherein the base unit includes a second microphone array.

Application No. 09/692,846
Filed 10/19/00
Page 45

40. (new) The system of Claim 39 wherein the first microphone comprises a first microphone array, and said first microphone array and the second microphone array implement a 3 dimensional microphone array.

41. (new) The system of Claim 26 further comprising a central database coupled to said external network interface, said central database including at least one of grammars; speech models; device abstractions; programming information; and lexica.

42. (new) The system of Claim 41 wherein said central database is coupled to said external network interface through an external network.

43. (new) The system of Claim 42 further comprising:
a remote server coupled to said external network and to said central database.

44. (new) The system of Claim 42 further comprising:

Application No. 09/692,846
Filed 10/19/00
Page 46

another natural language interface control system; and
another external network interface coupled to said
other natural language interface control system, and to said
external network.

45. (new) A natural language interface control system
for operating a plurality of devices comprising:

a first microphone;

a feature extraction, speech recognition and natural
language interface module coupled to the first microphone;

a device interface coupled to the feature extraction,
speech recognition and natural language interface module,
wherein the feature extraction, speech recognition and natural
language interface module is for operating at least one device
coupled to the device interface based upon non-prompted, open-
ended natural language requests; and

an external network interface coupled to the natural
language interface control system.

Application No. 09/692,846
Filed 10/19/00
Page 47

46. (new) The system of Claim 45 further comprising a central database coupled to said external network interface, said central database including at least one of additional grammars; additional hidden Markov models; additional device abstractions; programming information; and lexica.

47. (new) The system of Claim 46 wherein said central database is coupled to said external network interface through an external network.

48. (new) The system of Claim 47 further comprising:
a remote server coupled to said external network and to said central database.

49. (new) The system of Claim 47 further comprising:
another natural language interface control system; and
another external network interface coupled to said other natural language interface control system, and to said external network.

Application No. 09/692,846
Filed 10/19/00
Page 48

50. (new) A method of making a natural language interface control system for operating a plurality of devices comprising:

providing a first microphone;

coupling a feature extraction, speech recognition and natural language interface module to the first microphone;

coupling a device interface to the feature extraction, speech recognition and natural language interface module, wherein the feature extraction, speech recognition and natural language interface module is for operating at least one device coupled to the device interface based upon non-prompted, open-ended natural language requests; and

coupling an external network interface to the natural language interface control system.

51. (new) The method of Claim 50 further comprising; coupling a central database to said external network interface, said central database including at least one of additional grammars; additional hidden Markov models; additional device abstractions; programming information; and lexica.

Application No. 09/692,846
Filed 10/19/00
Page 49

52. (new) The method of Claim 51 further comprising coupling said central database to said external network interface through an external network.

53. (new) The method of Claim 52 further comprising: coupling a remote server to said external network and to said central database.

54. (new) The method of Claim 52 further comprising: providing another natural language interface control system; and

coupling another external network interface to said other natural language interface control system, and to said external network.

55. (new) A natural language interface control method comprising:

providing a feature extraction, speech recognition and natural language interface module;

Application No. 09/692,846
Filed 10/19/00
Page 50

coupling the feature extraction, speech recognition and natural language interface module to an external network; and

downloading at least one of grammars; speech models; device abstractions; programming information; and lexica into the feature extraction, speech recognition and natural language interface module through the external network.

56. (new) A natural language interface control method comprising:

providing a feature extraction, speech recognition and natural language interface module;

coupling the feature extraction, speech recognition and natural language interface module to an external network; and

transmitting at least one of grammars; speech models; device abstractions; programming information; and lexica from the feature extraction, speech recognition and natural language interface module to the external network.

Application No. 09/692,846
Filed 10/19/00
Page 51

Cancel claims 18 through 25 without prejudice or
disclaimer.